

Claims

*Sub A1*

1. A method for distributing requests for web pages among two or more web servers, comprising:  
monitoring the web servers to determine if a predetermined condition exists at one or more of the web servers; and  
if the predetermined condition does exist at least one of the web servers, transferring at least one request from that web server to another one of the web servers.

2. The method of claim 1 wherein the monitoring step comprises monitoring the web server's web page request queue length.

3. The method of claim 2 wherein the predetermined condition comprises a web page request queue length greater than a predetermined value.

4. The method of claim 1 wherein the monitoring step comprises monitoring the web server's web page requests queue delay.

5. The method of claim 1 wherein the predetermined condition comprises a web page request queue delay greater than a predetermined value.

*Sub D1*

6. The method of claim 1 wherein the monitoring step comprises monitoring the system load of the host running the web server.

7. The method of claim 6 wherein the predetermined condition comprises a CPU utilization greater than a predetermined value.

8. The method of claim 6 wherein the predetermined condition comprises a memory utilization greater than a predetermined value.

9. The method of claim 1 wherein the transferring step comprises transferring only if the request is for one of a predetermined set of web pages.

10. The method of claim 9 wherein the predetermined set is predetermined by a list of web pages included in the set.

11. The method of claim 9 wherein the predetermined set is predetermined by a list of web pages excluded from the set.

12. The method of claim 1 wherein the transferring step comprises transferring only if the request is for a web page that does not have state.

13. The method of claim 12 wherein the transferring step comprises:

determining whether the web page is included in a list web pages that have state;

transferring only if the web page is not included in the list.

14. The method of claim 1 wherein the monitoring step comprises monitoring the web servers to determine if the predetermined condition exists and wherein the predetermined condition comprises a failure.

15. A system for distributing requests for web pages, comprising:

a manager for monitoring web servers to determine if a predetermined condition exists at one or more of the web servers; and

a web server capable of transferring at least one request from the web server to another of the web servers if the predetermined condition does exist at least one of the web servers.

16. The system of claim 15 wherein the web server is capable of transferring only if the request is for one of a predetermined set of web pages.

17. The system of claim 15 further comprising an agent for determining the web server's web page request queue length and for transmitting the queue length to the manager.

Sub  
A2

18. The system of claim 15 further comprising an agent in communication with the manager for determining the web server's web page request queue delay and for transmitting the queue delay to the manager.

19. The system of claim 15 wherein the web server comprises a web server interface in communication with the web server, and wherein the web server interface directs the web server to transfer at least one request.

20. A method for distributing web page requests among two or more web servers, comprising:

periodically monitoring a web server load metric; and

transferring a request from one web server to another web server thereby balancing the load metric on each web server.

Sub  
00000000000000000000000000000000

Add B1

Add C1